# Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo

JOURNAL OF VIROLOGY, Nov. 1991, p. 6266–6276
0022-538X/91/116266-11$02.00/0

Vol. 65, No. 11

# Discontinuous Sequence Change of Human Immunodeficiency Virus (HIV) Type 1 *env* Sequences in Plasma Viral and Lymphocyte-Associated Proviral Populations In Vivo: Implications for Models of HIV Pathogenesis

PETER SIMMONDS,[1]* LIN QI ZHANG,[2] FIONA McOMISH,[2] PETER BALFE,[2] CHRISTOPHER A. LUDLAM,[3] AND ANDREW J. LEIGH BROWN[2]

*Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG,[1] Institute of Cell, Animal and Population Biology, Division of Biological Sciences, University of Edinburgh, Kings Buildings, Edinburgh, EH9 3JN,[2] and Department of Haematology, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh, EH3 9YW,[3] Scotland*

Received 10 June 1991/Accepted 13 August 1991

**Sequence change in different hypervariable regions of the external membrane glycoprotein (gp120) of human immunodeficiency virus type 1 (HIV-1) was studied. Viral RNA associated with cell-free virus particles circulating in plasma and proviral DNA present in HIV-infected peripheral blood mononuclear cells (PBMCs) were extracted from blood samples of two currently asymptomatic hemophiliac patients over a 5-year period. HIV sequences were amplified by polymerase chain reaction to allow analysis in the V3, V4, and V5 hypervariable regions of gp120. Rapid sequence change, consisting of regular replacements by a succession of distinct viral populations, was found in both plasma virus and PBMC provirus populations. Significant differences between the frequencies of sequence variants in DNA and RNA populations within the same sample were observed, indicating that at any one time point, the predominant plasma virus variants were antigenically distinct from viruses encoded by HIV DNA sequences in PBMCs. How these findings contribute to current models of HIV pathogenesis is discussed.**

The high degree of sequence variability that exists between different isolates of human immunodeficiency virus type 1 (HIV-1) (1, 38) poses a major problem for the development of effective methods of immunization against the virus. In particular, a major site for antibody-mediated virus neutralization in the *env* gene (the V3 hypervariable region [15, 17, 22, 25, 27]) shows considerable sequence heterogeneity and rapid rates of sequence change (1, 3, 12, 20, 36, 38, 43, 46). Furthermore, many of the amino acid changes in this region have been shown to modulate immunological recognition (22, 24, 30, 40).

We have used phylogenetic analysis of nucleotide sequences from a set of five serial samples from a (currently) asymptomatic hemophiliac patient infected with HIV-1 to investigate the rate and direction of sequence change in each of three hypervariable regions (V3, V4, and V5 [27]). By using a nested polymerase chain reaction (PCR) to amplify viral nucleic acids in vivo (37, 49), sequences of proviral DNA from peripheral blood mononuclear cells (PBMCs) were compared with those of viral RNA in plasma. We observed significant differences between the two populations in all three hypervariable regions at different points after infection. We present and discuss a model of HIV pathogenesis that takes these results into account along with the results of previous investigations of biological heterogeneity of HIV (2, 6, 41), the cell types infected with HIV in vivo (34, 35), and the evidence for positive selection for sequence change in hypervariable regions of the *env* gene (5, 36, 46).

* Corresponding author.

## MATERIALS AND METHODS

**Patient samples.** Sequential samples from a hemophiliac patient, p82, infected with HIV-1 from factor VIII prepared from Scottish blood donations in 1983 (23), were used for sequential studies of HIV sequence change. Seroconversion took place in June 1984, at which time a plasma sample was stored. Subsequent samples (from both plasma and PBMCs) from this patient were collected in June 1987, January 1988, February 1989 (1989A), and August 1989 (1989B). Several of the batches of factor VIII transfused to p82 were given to another hemophiliac patient, p80, who also seroconverted in 1984. A PBMC sample from this second patient was taken in February 1989 and was used for sequence comparisons.

**PCR product length analysis.** Sequence variants that differed in length in the V4 and V5 hypervariable regions were visualized by high-resolution gel electrophoresis of amplified DNA (36, 45). For V4 sequences, proviral DNA or cDNA was amplified first with primers e and h and then, in a second PCR, with primer f and a new antisense primer lying in the C3 region (5′ ATGGGAGGGCATACATTGC; position 7539 in pHIVHXB2). To amplify V5 sequences, the second PCR was carried out with primer g and a new sense primer in the C3 region (5′ GGAAAAGCAATGTATGCCC; position 7515 in pHIVHXB2). The relative proportions of sequence variants of different molecular weights within a sample was obtained by replicate amplification of undiluted proviral DNA (or viral cDNA) samples containing typically 100 to 200 molecules of target sequence. Bands were quantified by scanning densitometry of the autoradiographic image of the polyacrylamide gel with a Shimadzu densitometer.

**Sequencing of HIV gp120.** Proviral DNA was extracted from PBMCs as previously described (37). Single molecules of provirus were amplified after prior limiting dilution and

directly sequenced to avoid errors introduced by the amplification of DNA in vitro (3, 36). Viral RNA was extracted from plasma, reverse transcribed, and sequenced as previously described (49). The region amplified spanned the V3, V4, and V5 hypervariable regions (27) and was amplified with primers a to h in three sequential nested PCRs as previously described (36). This method gave continuous sequence from nucleotides equivalent to positions 6957 to 7814 in clone pHXB2 of $HIV_{HTLV\text{-}IIIB}$. Sequences from V3, V4, and V5 and surrounding nucleotides are presented in this article. A fuller analysis of the V3 sequences is in preparation.

**Evolutionary analysis of gp120 sequences.** Sequences obtained in this and our previous study were aligned by the Needleman and Wunsch algorithm as implemented by the program GAP on the University of Wisconsin GCG Package (8) and subsequently edited by hand. A matrix of evolutionary distances was generated by using the two-parameter model of Kimura (18); alignments that gave the minimum evolutionary distance between sequences were used in this study. Phylogenetic trees were constructed on the basis of the distance matrices by using the Fitch-Margoliash procedure (10) as available in program FITCH of the PHYLIP package supplied by J. Felsenstein. The validity of the trees was assessed by reentering the tree obtained by FITCH into the maximum-likelihood-based program DNAML as a user-defined tree. This gave confidence intervals for each of the internodal distances.

**Nucleotide sequence accession numbers.** Sequences obtained in this study have been submitted to GenBank under accession numbers M77541 through M77636.

## RESULTS

**Nucleotide sequence variation in V4 and V5.** A longitudinal study of HIV sequence change was carried out with samples from p82, a hemophiliac patient infected with HIV-1 from factor VIII concentrate. Sequences in the V4 and V5 regions were obtained from stored plasma from this individual at the time of acute seroconversion (May 1984) and subsequently from plasma and PBMCs at each of four time points following infection (June 1987, January 1988, February 1989 [1989A], and August 1989 [1989B]). This individual remained asymptomatic during the course of the study and has not received zidovudine or other antiviral therapy at any time. The results of standard virological and immunological investigations of this individual are shown in Table 1.

For comparative purposes, proviral DNA in a PBMC sample from another hemophiliac patient, p80, taken 5 years after infection, was also sequenced. This patient was infected by contaminated factor VIII around the same time as p82 (May 1984) and has also remained asymptomatic over the study period.

A total of 86 sequences in the V4 hypervariable region and 70 in the V5 region of *env* were obtained from p82; 9 V4 and 9 V5 sequences were obtained from the single PBMC sample from p80. The sequences obtained over the 5-year course of the longitudinal study were highly variable in the V4 and V5 regions, indicating rapid and continuous sequence change over the asymptomatic period of the infection. Each V4 and V5 nucleotide sequence was aligned against all the others by using a standard algorithm (GAP; see Materials and Methods), and a nucleotide distance matrix was obtained from pairwise comparisons. The evolutionary relationships between the different sequences were estimated by the FITCH program. In the resulting phenograms (Fig. 1a and b), evolutionary distances are shown by the horizontal separation between pairs of sequences (the vertical lines are of no significance).

TABLE 1. Standard virological and immunological markers of HIV infection in p82 and p80

| Sample | Time (months)[a] | $CD4^+$ lymphocytes ($10^9$) | p24 antigen[b] | Provirus-bearing PBMCs[c] |
|---|---|---|---|---|
| p82 | | | | |
| 1983 | −14 | 1.45 | NA | NA |
| 1984 | 0 | 0.93 | + | ND |
| 1987 | 36 | 0.53 | − | 1/2,000 |
| 1988 | 43 | 0.34 | − | 1/2,270 |
| 1989A | 56 | 0.65 | + | 1/700 |
| 1989B | 63 | 0.16 | − | 1/800 |
| p80 (1989) | 32 | 0.42 | − | 1/50,000 |

[a] Time from first positive serum sample.
[b] Detection of serum antigen by capture enzyme-linked immunosorbent assay (ELISA) (>15 pg/ml; Dupont). NA, not applicable.
[c] Proportion of PBMCs bearing provirus, estimated by limiting dilution (37). ND, not done.

A notable feature of this analysis is the apparent clustering of sequence variants into a small number of groups. In the V4 region (Fig. 1a), three groups can readily be identified (A, B, and C). Only one sequence, lying between groups A and C, does not fit into the classification. Clusters of distinct sequence types are also discernible in the V5 region (Fig. 1b), although in this case there are more groups (here labelled A to E) and some sequences that do not fit any of the groups (indicated by ?). Sequence variation within the V4 and V5 groups is considerably less than that which exists between groups. Unexpectedly, some of the sequences from p80 were identical to those of p82 in the V4 region (group A), while some had diverged to form a group clearly distinct from V4A, -B, or -C (Fig. 1a). Similarly, some of the V5 sequences from p80 were identical to those of p82 (V5A), while others fit none of the other p82 groups (Fig. 1b). As p80 and p82 shared several batches of noncommercial factor VIII in the year prior to seroconversion, and in view of the presence of identical V4 and V5 sequences in both, we infer that they were infected from the same source. Sequences of the V4A and V5A type are likely to have formed a major component of the virus population that infected both patients.

The mean within-group distances were 4.6, 5.6, and 1.0% for the V4A, -B, and -C groups, respectively, while the mean intergroup distances ranged from 11.0% (V4A to V4C) to 30.4% (V4B to V4C). In the V5 region, intergroup distances ranged from 16 to 55%, while within-group variability in no case exceeded 6.5%. In the V4 region, the branching pattern suggests that V4B and V4C sequence types diverged independently from V4A, the group that contains sequences found at seroconversion and those that are shared between p80 and p82. The major V5 groups also appear to have evolved independently from V5A. However, the distances between groups are so large that definite conclusions about such relationships cannot be made with these sequence data.

**Amino acid sequence variation in V4 and V5.** Figure 2 shows the translated sequences in the V4 region, divided into the groups indicated by the phenogram, to illustrate the differences between sequences within groups and the much greater differences between groups. The consensus sequences of V4A, -B, and -C are clearly distinct from each

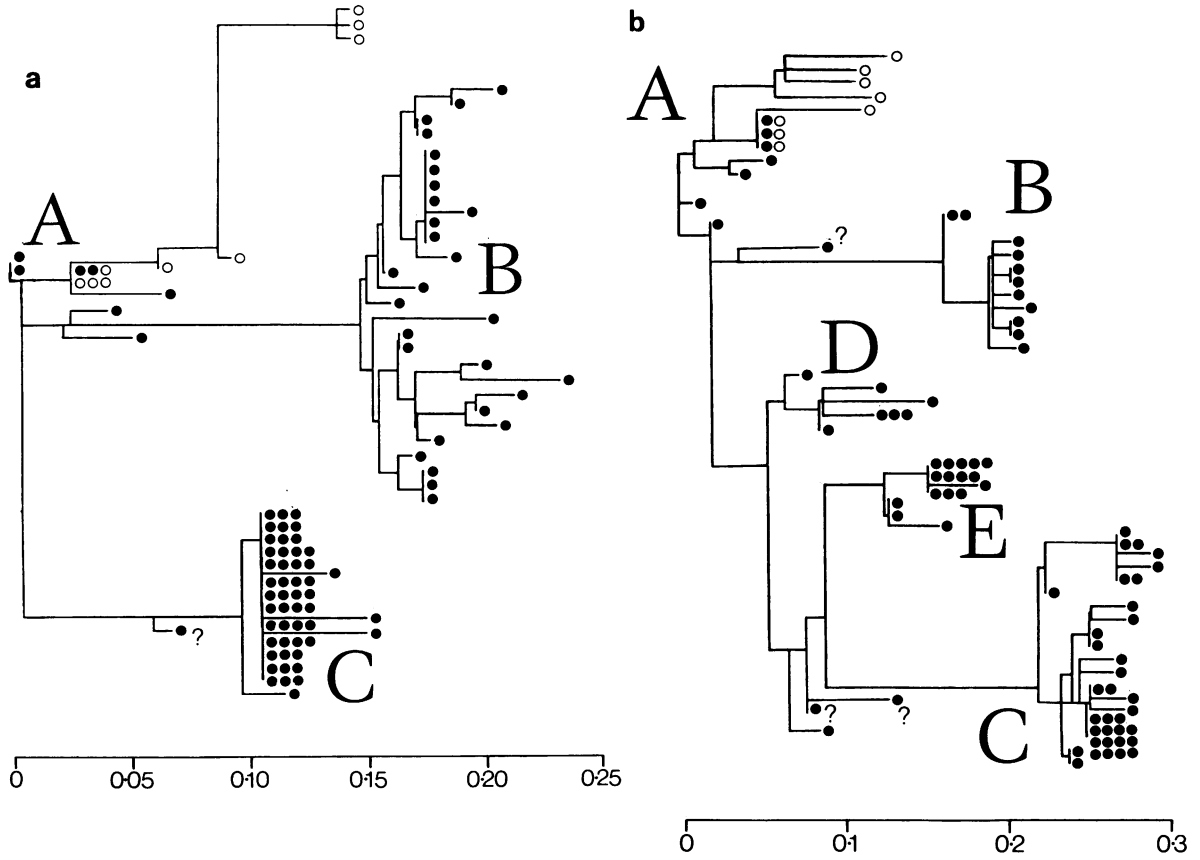


FIG. 1. Phylogenetic analysis of V4 (a) and V5 (b) nucleotide sequences from p80 (○) and p82 (●) over a 5-year period from the time of seroconversion. Sequence types are indicated as A to C (V4) or A to E (V5); intermediate and unclassified sequences are indicated by ?. Evolutionary distances between pairs of sequences are proportional to horizontal separation, as indicated on the scale. Maximum-likelihood analysis indicated that the distances between all nodes represented were significantly different from 0.

other, whereas in each of the three groups, individual sequences rarely differ from each other by more than two to three amino acid residues. Similarly, the potential sites for N-linked glycosylation in the hypervariable regions differ considerably between groups (indicated by # in Fig. 2). Within groups, the number and spacing of sites are relatively constant, although V4B is more variable in this respect, containing similar numbers of variants that differ at one or two of the four potential sites in the region.

The V4 region is known to show considerable variability in length among different isolates (1, 12, 27, 38) and also between proviral sequence variants within a single sample (36, 45). Comparison of the consensus sequences in this region from p82 (Fig. 2) reveals that virtually all V4B variants consist of 17 amino acid residues between the relatively more-conserved flanking regions (..FNSTW----<V4>----ITLPCR...), with only three exceptions (sequences 7, 8, and 16) which are two or three amino acids longer. Similarly, the lengths of all but three (2, 3, and 4) of the V4C sequences are constant, at 18 amino acids between the conserved flanking regions.

The lengths of the V4A sequences are somewhat more variable. Sequences of type 1 of V4A from p82 and the four sequences of type 1 of p80 are identical at both the amino acid and nucleotide levels and are all 15 amino acids long, the same length as V4A-2 and V4A-3. However, in both individuals, longer V4 sequences are also found: in p82, there were three additional sequences of 20, 22, and 24 residues (sequences 5, 4, and 3 respectively), and in p80, there were four sequences of 27 amino acids and one of 26.

Variation in the V5 region (Fig. 3) shows many of the characteristics of variation seen in V4. The consensus sequences of the five groups (V5A to E) differ considerably from one another, while sequence variation within groups is minimal, particularly in groups B to E. Group A sequences, defined by the phylogenetic tree (see above), appear to contain two types of sequences at the amino acid level (sequences 4 and 5 appear distinct from the others), although for the purposes of analysis (see below), the numbers of sequences are so small as not to justify further subgrouping. The pattern of N-linked glycosylation sites is also well conserved within groups, and the overall lengths of the regions (between ...TRDGG---<V5>---FRPGG....) are 7 to 10, 8, 12, 8, and 8 residues in groups A to E, respectively. Sequence type 1 from p80 ($n$ = 3) in the V5 region is identical to V5A-1 ($n$ = 3) of p82 (Fig. 3), while the other relatively small number of other sequences from p80 differ considerably from the common type and from each other. Further analysis would be necessary to find out whether sequences from this patient grouped into distinct types as they appear to do in this region from p82.

**Sequence change in the V4 and V5 regions.** Having defined and analyzed the sequence groups in the two hypervariable regions, the classification can be used to study sequence

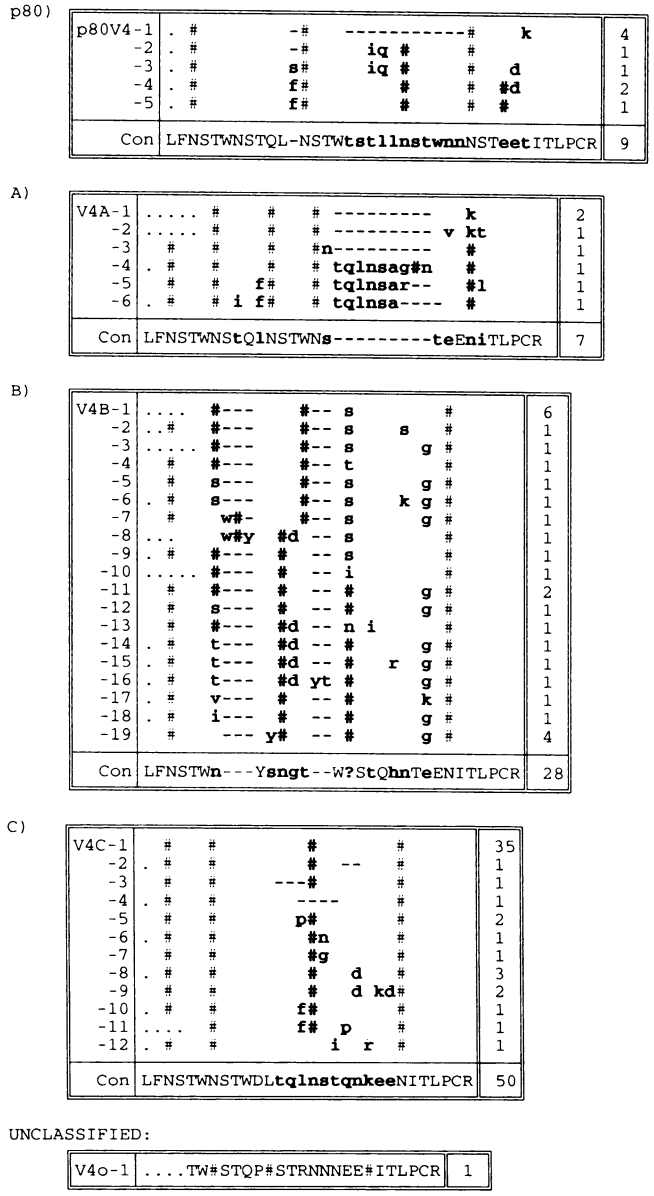


FIG. 2. Peptide sequences of the three phylogenetic groups in the V4 hypervariable region. Con, consensus sequence for each group; nonconserved residues are shown in boldface lowercase letters. Differences from consensus are shown for each sequence entry, and frequency of detection of each sequence type is shown in the rightmost column. Symbols: ?, no majority at this position;., unsequenced; –, gap introduced to preserved sequence alignment within group; #, all potential sites of N-linked glycosylation (nonconserved sites shown in boldface).

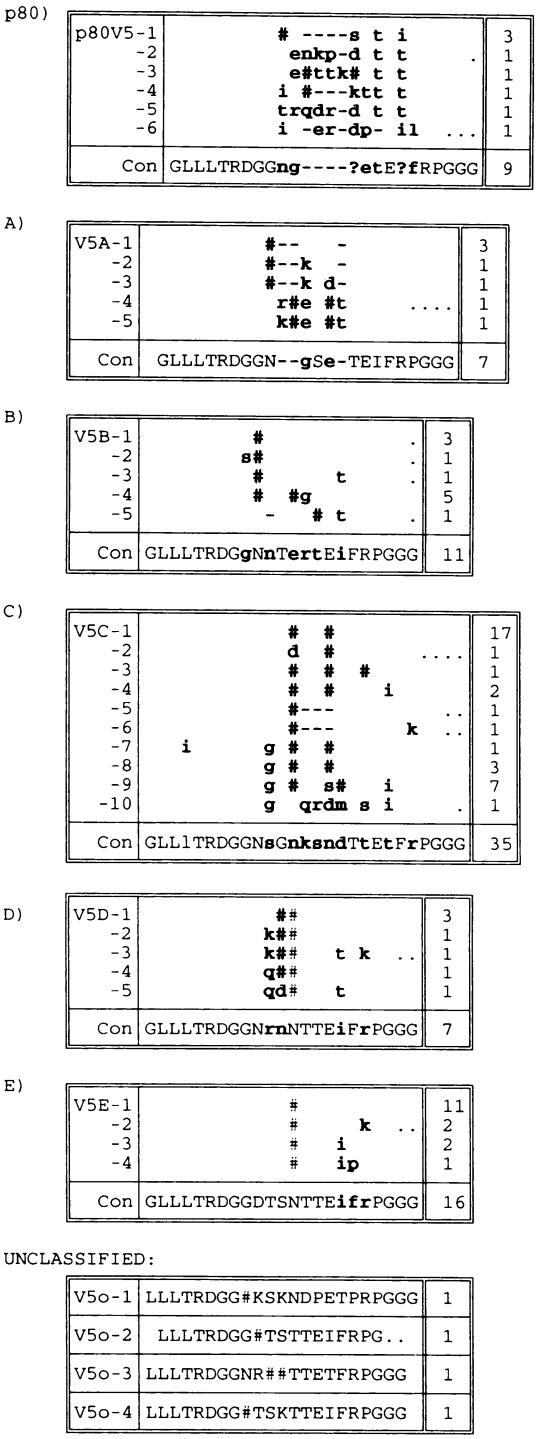


FIG. 3. Peptide sequences of the five phylogenetic groups in the V5 hypervariable region. Arrangement and symbols are same as for Fig. 2.

changes over time in samples from p82. Figure 4a records the numbers of sequences detected in plasma (above the *x* axis), and PBMCs (below the *x* axis). Although the numbers of sequences obtained at any one time point are relatively low, there is clear evidence for turnover of sequence variants. Type A variants were found in all four of the seroconversion RNA sequences in 1984, while only sequences of type B were found in 1987. In the following year, the most commonly found sequence was type C, which appears to have completely replaced type B in the two samples taken in 1989. Turnover of sequence variants may also be observed in the provirus population. Both V4A and V4B sequences were found in 1987, while in the following year almost all sequences were of type B. The replacement of V4B with V4C was completed in the following year.

Comparable turnover of sequence variants is also found in

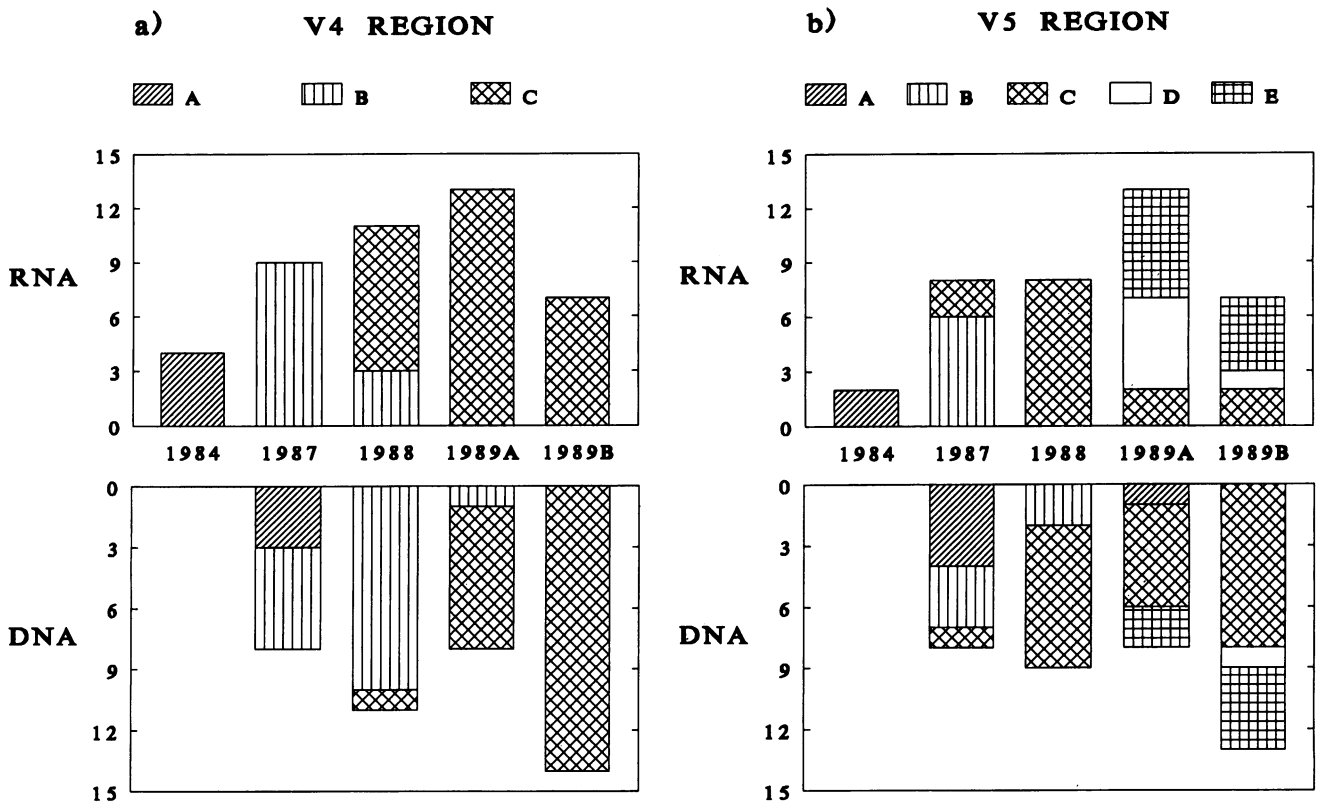


FIG. 4. Frequency of detection of V4 (a) and V5 (b) sequence types in sequential samples from p82 (1984 to 1989B). RNA sequences are shown above the $x$ axis, and DNA sequences are shown below the $x$ axis.

the V5 hypervariable region (Fig. 4b). Whereas V5A sequences were found at seroconversion in the plasma, these were replaced successively by V5B in 1987, V5C in 1988, and finally by V5D and -E in 1989. A similar progression was also observed in the proviral population. In 1987, approximately half of the sequences were of seroconversion type A. The almost complete replacement of these sequences by V5B and V5C took place in the following two years. In turn, V5D and -E appeared to be in the process of replacing V5C by the end of the study period.

**Differences between DNA (proviral) and RNA (viral) populations.** At several time points, there were considerable differences in the frequencies of different sequence types in the DNA and RNA samples. For example, the 1988 DNA sample contained predominantly V4B sequences in PBMCs (10 of 11) yet mainly V4C sequences in the plasma samples (8 of 11). Similarly, the preponderance of V5C sequences in the two 1989 DNA samples (6 of 9 and 8 of 13) contrasted with the infrequency of their detection in the corresponding plasma samples (2 of 13 and 2 of 7). The appropriate statistical procedure for comparing frequencies in small samples, Fisher's exact test, was used to test the significance of these differences. The frequencies of V4 variants in the 1988 sample and of V5 variants in the 1989A sample were found to be significantly different between the PBMC proviral and plasma RNA populations ($P < 0.01$ and $P < 0.05$, respectively).

The relative frequencies of the various sequence types in the V4 and V5 regions was also estimated by high-resolution gel electrophoresis of amplified DNA (36, 45). As indicated previously (Fig. 2), 25 of the 28 V4B sequences had an overall length of 17 amino acids, while 47 of the 50 V4C sequences were 18 amino acids long. Aliquots of DNA (2 μg) extracted from the PBMC samples in 1988 and 1989A containing approximately 70 and 220 molecules of provirus (Table 1) and undiluted cDNA after reverse transcription of RNA sequences present in plasma (containing 100 to 200 copies of target sequence; data not shown) were amplified in two stages with primers specific for the V4 region (see Materials and Methods). The product DNA consisted of two size variants, differing in electrophoretic mobility by 3 bp (Fig. 5). As indicated, the smaller band corresponds to the predicted size of V4B and the larger band corresponds to V4C. The 1988 DNA sample (lanes b) consists of mainly V4B sequences, while the corresponding RNA sample (lanes c) consists of predominantly V4C. The almost complete replacement of V4B sequences by V4C between 1988 and 1989 (Fig. 4) is also shown by this length analysis; both DNA (lanes d) and RNA (lanes e) contain predominantly V4C sequence types.

The relative numbers of V4B and V4C sequences were quantified by scanning densitometry (Table 2). To show that representative numbers of sequence variants had been amplified, each sample was amplified in replicate to allow two independent samplings of the populations present. The reasonably close agreement between all of the replicates confirmed that the populations studied (>100 sequences in each sample) were sufficiently large to prove that the differences between the populations at both time points were not due to sampling error. Furthermore, the relative proportions agree closely with those determined by sequence analysis (Fig. 4). For example, the 1988 DNA sample contained 74 to 75%

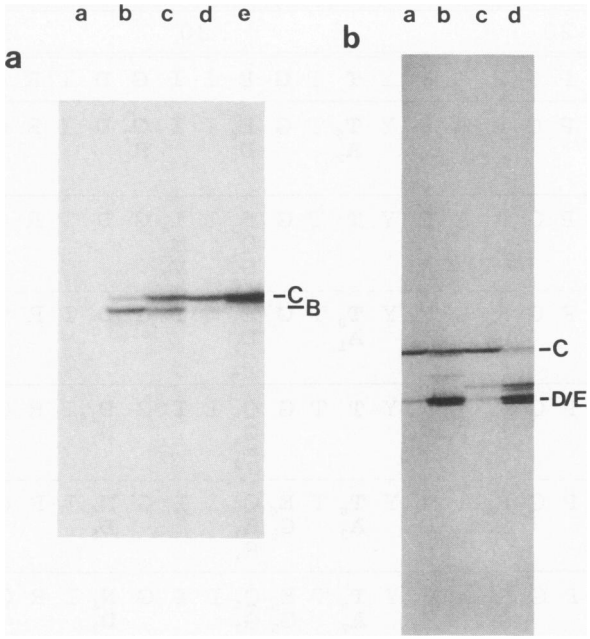


FIG. 5. Length analysis of amplified DNA in the V4 (a) and V5 (b) hypervariable regions to confirm existence of population differences in the in vivo DNA and RNA populations. (a) Lanes: a, negative human DNA amplified with primers spanning the V4 region; b and d, PCR product from V4 region of proviral DNA from the 1988 and 1989A samples, respectively, from p82; c and e, PCR product from viral RNA in the corresponding plasma samples. Expected sizes of V4B and V4C sequences are indicated. (b) Lanes: a and c, PCR product from V5 region of proviral DNA from the 1989A and 1989B samples, respectively; b and d, corresponding RNA samples. Expected sizes of V4C and V4D and -E are indicated.

V4B sequences while the RNA sample contained only 42 to 48%. The corresponding numbers of V4B sequences are 10 of 11 and 3 of 11 in these two samples. Similarly, the 1989 DNA sample contained 84 to 85% V4C sequences by densitometry, compared with 10 of 11 by sequence analysis, while the RNA population was uniformly V4C by both methods.

TABLE 2. Serial changes in the frequencies of V4 and V5 length variants in plasma (RNA) and PBMC (DNA) samples from p82 estimated by densitometry

| Sample | Year[a] | Variant[b] | | | | | |
|---|---|---|---|---|---|---|---|
| | | V4B (%) | V4C (%) | Ratio | V5C (%) | V5D, -E (%) | Ratio |
| DNA | 1988 | 75, 74 | 25, 26 | 2.92 | ND[c] | ND | ND |
| RNA | 1988 | 48, 42 | 52, 58 | 0.82 | ND | ND | ND |
| DNA | 1989A | 15, 16 | 85, 84 | 0.18 | 77, 63 | 23, 37 | 2.33 |
| RNA | 1989A | 0, 0 | 100, 100 | 0.00 | 32 | 68 | 0.47 |
| DNA | 1989B | ND | ND | ND | 75, 78 | 25, 22 | 3.26 |
| RNA | 1989B | ND | ND | ND | 18 | 82 | 0.22 |

[a] 1989A, samples were taken in February 1989; 1989B, samples were taken in August 1989.
[b] Percents are replicate densities of independently amplified aliquots of the original DNA or cDNA. Repeat samples were not available from RNA samples in the V5 region.
[c] ND, not done.

TABLE 3. Frequencies of the combinations of V4 and V5 types in the 75 complete sequences obtained in the study

| V5 sequence type | V4 sequence type | | |
|---|---|---|---|
| | A | B | C |
| A | 5 | 2 | |
| B | | 11 | |
| C | | 9 | 25 |
| D | | | 7 |
| E | | | 16 |

An equivalent analysis of the V5 region was carried out, and the discrepancy between the relative numbers of (i) V5C and (ii) V5D and V5E in the two populations was investigated. V5C differs in length from V5D and -E by 12 nucleotides (Fig. 3). Figure 5b confirms the existence of a marked difference in the relative numbers of the two sequence types in both the 1989A and 1989B samples. Ignoring the sequences that are of intermediate length between the two main types and whose classification is uncertain, the majority of DNA sequences are V5C in both the 1989A and 1989B samples (63 to 78%), which is comparable to the numbers of sequences found previously (6 of 9 and 8 of 13). By contrast, 68 to 82% of the corresponding RNA sequences were of type D or E, reflecting their frequency of detection by sequence analysis (11 of 13, 5 of 7) at the two time points.

Having established by two methods that significant differences exist between the two populations at at least three time points, a more detailed consideration of the origin of these differences is justified. A general trend that is found in both the V4 and V5 regions is for RNA sequences to turn over more rapidly than the corresponding DNA sequences (Fig. 3). For example, the seroconversion type V4A sequence is completely replaced in plasma by 1987 yet forms a substantial proportion of sequences in PBMCs at that time. Similarly, the difference in the relative numbers of V4B and -C sequences in the 1988 sample and the replacement of V5C with the V5D and -E variants in the 1989 sample could be interpreted as a more rapid transition to a new sequence type in the plasma. The possible mechanisms and the consequences of this observation are discussed below in relation to current models of HIV pathogenesis.

**Linkage of V4 and V5 sequences.** To obtain the sequences in this study, single molecules of HIV provirus or RNA reverse transcript were isolated by limiting dilution prior to amplification with primers spanning the entire V3-C2-V4-C3-V5 region. With this method, we have obtained sequences that are not only free of errors associated with copying of DNA in vitro but also have avoided the problems of producing sequences that are hybrids of two or more HIV sequences present in the patient sample because of switching between different templates during the amplification process (26). These sequences can therefore be used for studies of linkage and recombination in vivo.

A very restricted number of combinations of V4 and V5 sequences were observed in our datum set (Table 3). We found that there was no fixed relationship between a given V4 sequence type with those of V5. For example, HIV sequences of type V4A could contain either V5A or V5B sequences; similarly, V4B was associated with V5A, -B, and -C. Finally, as well as being linked to V5C, V4C was also found in viral sequences containing the V5D and V5E sequence types. A consequence of the variable associations

| SAMPLE | n | 10 | | | | | | | | | | 20 | | | | | | | | | | 30 | | | | | | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1984 Plasma | 1 | N | N | T | R | K | **S** | I | **H** | I | G | P | G | **R** | **A** | **F** | Y | **T** | T | **G** | **E** | I | **I** | **G** | **D** | I | R | Q |
| Patient #80 (1989 PBMC) | 8 | N | N | T | R | K | **S** | I | **H**$_4$ **P**$_2$ **N**$_2$ | I | G | P | G | **R** | **A** | **F** | Y | **T**$_6$ **A**$_2$ | T | **G** | **E**$_6$ **D**$_2$ | I | **I** | **G**$_7$ **R**$_1$ | **D** | I | R | Q |
| 1987 PBMC | 5 | N | N | T | R | K | **S** | I | **H**$_4$ **P**$_1$ | I | G | P | G | **R** | **A** | **F** | Y | **T** | T | **G** | **E**$_3$ **Q**$_1$ **G**$_1$ | I | **I**$_3$ **M**$_1$ **V**$_1$ | **G** | **D** | I | R | Q |
| 1987 Plasma | 9 | N | N | T | R | K | **S**$_8$ **G**$_1$ | I | **H**$_5$ **P**$_4$ | I | G | P | G | **R**$_8$ **G**$_1$ **S**$_1$ | **A** | **F** | Y | **T**$_8$ **A**$_1$ | T | **G** | **E**$_3$ **D**$_3$ **G**$_2$ | I | **I** | **G** | **D** | I | R | Q |
| 1988 PBMC | 11 | N | N | T | R | K | **S**$_7$ **R**$_3$ **G**$_1$ | I | **H**$_8$ **P**$_3$ | I | G | P | G | **R**$_9$ **S**$_1$ **G**$_1$ | **A**$_{10}$ **T**$_1$ | **F**$_8$ **V**$_3$ | Y | **T** | T | **G** | **Q**$_5$ **G**$_4$ **D**$_2$ | I | **I** | **G** | **D**$_{10}$ **N**$_1$ | I | R | Q |
| 1988 Plasma | 11 | N | N | T | R | K | **R**$_5$ **G**$_4$ **S**$_2$ | I | **H** | I | G | P | G | **R**$_8$ **S**$_3$ | **A** | **V**$_7$ **F**$_4$ | Y | **T**$_8$ **A**$_3$ | T | **E**$_6$ **G**$_5$ | **Q**$_7$ **G**$_3$ **R**$_1$ | I | **I** | **G** | **N**$_6$ **D**$_5$ | I | R | Q |
| 1989A PBMC | 8 | N | N | T | R | K | **R**$_6$ **G**$_2$ | I | **H**$_4$ **Y**$_4$ | I | G | P | G | **R**$_6$ **S**$_2$ | **A** | **V**$_6$ **F**$_2$ | Y | **T**$_6$ **A**$_2$ | T | **E**$_5$ **G**$_3$ | **Q**$_6$ **G**$_2$ | I | **I** | **G** | **N**$_4$ **D**$_3$ | I | R | Q |
| 1989A Plasma | 13 | N | N | T | R | K | **G**$_{11}$ **R**$_2$ | I | **H**$_{11}$ **Y**$_2$ | I | G | P | G | **S**$_{12}$ **R**$_1$ | **A** | **F**$_{10}$ **V**$_3$ | Y | **A**$_{11}$ **T**$_2$ **A**$_1$ | T | **G**$_{11}$ **E**$_2$ | **G**$_{10}$ **Q**$_2$ | I | **I** | **G** | **D**$_{11}$ **N**$_2$ | I | R | Q |
| HIV-MN | | Y | N | K | R | K | R | I | H | I | G | P | G | R | A | F | Y | T | T | K | N | I | I | G | T | I | R | Q |

FIG. 6. Sequences at the center of the V3 loop in sequential samples from p82 and a single DNA sample from p80 (sequence of $HIV_{MN}$ included for comparison). Numbering begins from the cysteine residue at the start of the V3 loop. Variable residues are shown in boldface type, numbers of the major and minor variants at variable sites are shown in subscript, and numbers of sequences obtained from each sample are indicated (n).

between hypervariable regions was that the frequencies of sequence types varied independently from each other. For example, it can be seen from Fig. 4 that the predominant virus type in the V4 region remained V4C at a time when V5 sequences were changing from V5C to V5D and -E. Similarly, while V4A sequences were being replaced by V4B in 1987 to 1988, V5 sequences underwent two replacements from V5A to V5B and then to V5C over the same interval. However, the changes in the frequencies precluded a statistical investigation of association between variants (linkage disequilibrium).

Combinations of V4 and V5 sequences showed a higher rate of turnover than that of the different sequences considered separately. This led to even greater differences between the DNA and RNA populations at a given time point. For example, in 1987, most of the RNA sequences were of combined type BB, while the DNA was predominantly BA. In 1988, RNA genotypes were almost exclusively CC, while those of DNA were mainly BC (data not shown). Significant differences between the frequencies of V4-V5 combinations between the DNA and RNA populations were found at all four time points (data not shown).

By using the data on the frequencies of V4 and V5 combinations, the following succession of genotypes was observed over the 5-year follow-up period:

$$AA \rightarrow BA \rightarrow BB \rightarrow BC \rightarrow CC \rightarrow CD$$
$$\searrow ?$$
$$\rightarrow CE$$

This temporal succession of genotypes does not necessarily imply that each preceding form was ancestral to the variant that succeeded it. In the V4 region (Fig. 1), V4B and V4C appear as lineages entirely independent from V4A, although the accuracy of the phylogenetic analysis is limited by the extremely high rate of sequence change in this region. If the derived V4 and V5 sequences have evolved independently from the seroconversion type, then recombination must have occurred in vivo to generate the combinations of sequences found (Table 3).

**Sequence change in the V3 region.** The role of the V4 and V5 regions in antigenic recognition has not yet been defined. To investigate whether the difference in the sequence types in the DNA and RNA populations would lead to alterations in the susceptibility of the virus to antibody-mediated neutralization, a set of sequences similar to those of the V4 region were obtained in V3. In Fig. 6, we show the sequences obtained from these samples at the center of the V3 loop which include epitopes that have been implicated in both antibody-mediated and cytotoxic T-cell recognition (22, 40). The single sequence obtained from the plasma sample from p82 at seroconversion differed from that of $HIV_{MN}$ at many of the sites shown. This sequence was identical to three of the five sequences in the 1987 DNA sample from p82 and to four of the eight sequences in the 1989 sample from p80. It is therefore likely to have formed a major component of the infectious virus population in the factor VIII given to both patients.

While the 1987 DNA and RNA samples were similar in V3 sequences, considerable differences between the DNA and

RNA samples in the frequencies of amino acids were observed in both the 1988 and 1989 samples from p82. For example, at residue 10 of the V3 loop, the majority of DNA sequences had an arginine in the 1989 sample, whereas the corresponding RNA sequences were generally glycine. Similar discrepancies were found at residues 17, 19, 21, 23, and 24. Substitutions at many of these residues have been previously shown to abolish serological or T-cell reactivity. Thus, most of the viruses encoded by the RNA sequences are probably quite different antigenically from those encoded by proviral sequences in the PBMCs. The significance of this finding for sequential studies of virus neutralization is discussed below.

## DISCUSSION

**Rate of sequence change in gp120.** In this study, we have produced evidence for rapid sequence change in three hypervariable regions of *env*. This finding was anticipated by our own cross-sectional studies of sequence evolution in a cohort of hemophiliac patients infected from a common source (3, 36) and of the V3 sequences of six children infected from a single plasma donation (46). Although neither study determined the infecting sequence population, the existence of substantial sequence variation among individuals 3 to 5 years after infection allowed an estimate of the rate of sequence change from a calculated common ancestor (in terms of percent nucleotides per year) to be made (3). The model used for this calculation allows for differences in the rate of sequence change among individuals, and the nucleotide distance estimates are corrected for multiple substitutions (18), but this approach does not account for convergence of sequences due to selection. It also assumes a steady accumulation of substitutions with time.

We have shown here that sequence evolution in p82 was indeed more rapid than that in p80, but, more important, that substitutions do not accumulate steadily with time. Sequence change in p82 over the five years of follow-up consisted of a series of replacements of one particular sequence type with another. We have shown that succeeding sequence types were not necessarily directly derived from the previous sequence; for example, V4C succeeded V4B in 1988 to 1989, yet V4B may not be the immediate ancestor of any of the V4C sequences (Fig. 1).

That evolution of HIV in vivo can be discontinuous is shown by the failure to detect intermediate forms between the major sequence types, despite the fact that numerous base changes and more than one insertion or deletion event have occurred in the development of variant V4 and V5 sequences from the seroconversion type. Further evidence for the existence of hidden evolution is provided by the repeated observations in both the V4 and V5 regions that each succeeding sequence type is not obviously more related to those that come before or after it than they are to the sequence of the original infecting virus.

It has frequently been argued that sequence change in the *env* region may be an adaptive response by HIV to evade recognition by the immune system. Several studies have shown high rates of amino acid substitutions precisely in those areas of the immunodominant loop that are the targets of B-cell and T-cell recognition (1, 3, 12, 20, 36, 38, 44, 46) and in the equivalent region of the simian immunodeficiency virus genome of infection of rhesus monkeys (5). Indirect evidence for positive selection for sequence change in V3 is provided by a depressed synonymous-to-nonsynonymous ratio ($K_s/K_a$ [21]) of nucleotide substitutions, significantly below 1, in the V3 loop region (36). In the current study, we have also found high rates of sequence change in these areas and could interpret the turnover of V3 sequence variants as a succession of escape mutants whose evolution is favored by a transient failure by the host immune system to neutralize the newly emergent forms. As was found in the V4 and V5 regions, succeeding V3 sequences are not necessarily direct derivatives of the previous V3 types. The predominant sequence in the 1989 RNA sample differs less from the seroconversion sequence than it does from the preceding variant (found at high frequency in the 1988 RNA sample and in the DNA population of the 1989 sample). As argued previously, whether the high rate of sequence change in the V4 and V5 regions is also a consequence of immune selection is not clear. Mouse antiserum to a peptide corresponding to the V5 region could neutralize HIV (15), although titers were lower than that of the antiserum raised against the V3 peptide, consistent with other peptide mapping studies (11, 17, 25, 28, 30, 32). However, the mature HIV gp120 protein is heavily glycosylated (19), and many of the potential sites for N-linked addition of carbohydrate are concentrated in the V4 and V5 hypervariable regions. These posttranslational modifications and long-range interactions with other regions of *env* on folding of the mature protein are likely to contribute to the formation of predominantly conformational epitopes in these regions.

**Long-term persistence of seroconversion-type sequences.** Many of the proviral sequences from p80 and p82 in samples taken several years after primary infection were identical to those detected at seroconversion. The absence of sequence change in the some of the most variable areas of the HIV genome is, at first sight, inconsistent with the generally high mutation rates associated with HIV replication (3, 33, 46). One explanation for complete absence of either silent or nonsilent changes over the entire V3-C2-V4-C3-V5 region is that the HIV proviral sequences detected in the PBMC samples in 1987 (p82) or 1989 (p80) that correspond to the seroconversion-type sequences have not replicated to any significant extent during the intervening years.

Supporting this hypothesis is the observation that HIV preferentially infects a long-lived cell subset of PBMCs in vivo. Almost all of the provirus detected in PBMCs is present in the $CD4^+$ lymphocyte fraction (35), of which the memory cell subset ($CD45RO^+$; $CD29^+$) appears to be preferentially infected in vivo by both HIV (34) and simian immunodeficiency virus (43). Consistent with their function in antigenic recall, it has been shown in humans (7) and by adoptive transfer in mice (16) that T memory cells can have essentially unlimited life spans relative to that of the host. Although HIV is normally considered cytopathic for T lymphocytes, a proportion of activated T lymphocytes may survive infection during the primary HIV infection and continue to circulate as differentiated memory cells with an unchanged proviral sequence. Thus, the persistence in p82 of V4B DNA sequences until 1988, when almost all RNA sequences were of the V4C type, may have been the consequence of long-term persistence of cells nonproductively infected in 1987.

There are many possible explanations for the proposed long-term survival of T cells infected at seroconversion and in subsequent years. Firstly, proviral sequences in those PBMCs that survive infection may contain inactivating mutations that prevent subsequent virus replication. High frequencies of defective proviral sequences have been reported to exist in vivo (26). However, using the limiting dilution PCR method that eliminates in vitro copying errors during

amplification (37), we have found an extremely low rate of inactivating substitutions in the *gag* and *env* regions (1 in over 40,000 bp sequenced) [3] and 3 in 110,000 bp [unpublished observations]). Furthermore, it has been shown that a high proportion of proviral sequences present in PBMCs can be activated in vitro to give replication-competent virus (4). Thus, defective viruses probably contribute little to persistent, nonlytic infection of lymphocytes.

An alternative explanation for the failure of HIV to kill the cell it infects is that the provirus may integrate into sites in the human genome that preclude or reduce the efficiency of cellular and viral mechanisms of transcription initiation, as has been described for other retroviruses (42).

Finally, the variants detected in the PBMC population may be replication competent and capable of activation but may contain mutations that make them less cytopathic for lymphocytes and allow the infected cells to survive. Supporting this latter possibility is the extensive in vitro evidence that isolates from HIV-infected individuals taken in early stages of infection are often noncytopathic for T lymphocytes, grow poorly in culture, and are incapable of any growth in T cell lines ("slow/low variants") (2, 6, 41). Virus variants in cells that have survived infection with HIV may therefore be a highly selected subset of the original infecting virus strain, whose noncytopathic (or defective) properties ensure their long-term survival without any need for continuous replication.

**Origin of plasma and PBMC virus populations.** We have provided extensive evidence for the existence of differences in the frequencies of different sequence types of virus present in plasma compared with those of proviruses present in HIV-infected PBMCs. Corroboration of the results obtained from sequencing (Fig. 3) was obtained by length analysis of amplified PCR products (Fig. 5 and Table 2), which discounted any effect of sampling error due to the small number of sequences. That there should be a difference between the two populations was unexpected, although it is not necessarily inconsistent with current theories of the pathogenesis of HIV (see below).

A consistent observation of this study was that each of the sequence types that initially appeared in the plasma RNA population eventually became the predominant PBMC sequence type. For this reason, the hypothesis that the unequal distribution of sequence types can be explained by their differing cell tropisms cannot be sustained. Similarly, comparisons of V3 proviral sequences in brain and spleen biopsy samples from three HIV-infected individuals has failed to reveal any systematic tissue-specific sequence differences (9). Only one of the three individuals showed major differences in the frequencies of distinct sequence types between the two tissues, and in the light of the data presented here, it is clearly possible that this difference was merely temporal.

The source of virus in the plasma could therefore be a subset of transcriptionally active $CD4^+$ lymphocytes, or virus could be secreted into the circulation by cells sequestered in solid tissue. It has been shown that plasma of both symptomatic and asymptomatic individuals is infectious (14), and thus infection of PBMCs may be a self-sustaining process. Infection, and continued sequence evolution of HIV, may indeed take place in peripheral $CD4^+$ lymphocytes.

Despite the high titers of infectivity of plasma in vitro, only a low proportion of T lymphocytes are infected in vivo (31, 35, 37), and an even lower proportion expresses detectable levels of virally encoded mRNA (13). Productive infection of T lymphocytes is thought to require T-cell activation by specific antigen or mitogen (39, 47, 48); thus, the observation that only 1 in 100 to 1 in 100,000 T lymphocytes contains provirus reflects the low frequency of activated cells in peripheral circulation. Although there is some evidence that the block to complete replication in nonactivated lymphocytes is at the level of integration and virus expression (39), it has been recently shown that virus replication may be prevented by incomplete reverse transcription of the incoming viral RNA (47). Furthermore, the truncated transcript is unstable and rapidly degraded, helping to explain the low frequency of provirus-bearing PBMCs in vivo.

According to the model advanced here, at any one time, proviral DNA sequences are composed of two distinct populations. Firstly, there are complete integrated copies of provirus in $CD45^+$ lymphocytes with no or minimal virus expression. Within the same sample, there would also be $CD4^+$ lymphocytes containing proviral DNA that were actively infected with HIV of the sequence types present in the plasma. The relative proportions of the two types of DNA would depend on the degree of infectivity of the plasma. We have previously shown that the amount of viral RNA present in plasma of HIV-infected individuals varies considerably, although there is a trend for symptomatic individuals to show higher concentrations than asymptomatic individuals (49). Similarly, the infectivity titers of plasma samples from patients with AIDS and AIDS-related complex were considerably higher than in those from patients with no evidence of clinical progression (14). Thus, in the early asymptomatic stages of infection, the majority of DNA sequences may remain of the seroconversion type, while, on progression, higher levels of infectious virus lead to increasing numbers of proviral sequences from secondary infection of lymphocytes, whose sequences would correspond to those of the plasma virus. It is notable that the apparent replacement of HIV sequences in PBMCs took place at a time when the proportion of infected cells was increasing (Table 1). Entirely consistent with the data given is the hypothesis that PBMCs containing the seroconversion type sequences remained in similar numbers for several years but were not detected after 1987, because they were numerically overtaken by the increasing numbers of PBMCs containing proviral sequences of the derived V4 and V5 sequence types.

The most direct test of whether plasma viral sequences are preferentially expressed in PBMCs is to compare the population of HIV mRNA sequences with those of PBMC provirus and plasma sequences. Unfortunately, it would not be possible to differentiate genuine mRNA sequences from viral RNA sequences that are present in the cytoplasm as a result of infection from plasma. According to the model advanced by Zack et al. (47), even nonactivated T cells may be susceptible to HIV attachment and entry; thus, within a PBMC sample, a large proportion of T lymphocytes may contain intact RNA templates from exogenous virus.

A major consequence of the difference in the compositions of the DNA and RNA populations is that sequential studies of virus evolution that are based on viral isolations from PBMC samples may be misleading. Because new sequence variants are initially more common in plasma than they are in PBMCs, isolations from the latter source may be composed predominantly or exclusively of previous virus types, possibly even of the seroconversion sequence. We are currently studying antigenic recognition of sequence-dependent epitopes in the V3 region to investigate the time course of development of specific immunity in p82, using oligopeptides

corresponding to each of the different RNA and DNA sequences obtained in this study.

## ACKNOWLEDGMENTS

We are indebted to A. Cleland for technical assistance, to the staff of the Haemophilia Centre, Royal Infirmary of Edinburgh, for collection of patient samples, and to S. Rebus, D. Beatson, J. F. Peutherer, and M. Steel for providing background virological and immunological data on the two hemophiliac patients studied. We acknowledge the assistance of J. McKeating, J. O. Bishop, E. Holmes, and H. G. Watson for critical reading of the manuscript.

The work was supported by the Medical Research Council AIDS Directed Programme. L. Q. Zhang is supported by grants from the British Council and from the People's Republic of China.

## REFERENCES

1. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63–74.
2. **Asjo, B., J. Albert, A. Karlsson, L. Morfeldt-Manson, G. Biberfeld, K. Lidman, and E. M. Fenyo.** 1986. Replicative properties of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet **ii**:660–662.
3. **Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating substitutions. J. Virol. **64**:6221–6233.
4. **Brinchmann, J. E., J. Albert, and F. Vartdal.** 1991. Few infected $CD4^+$ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. J. Virol. **65**:2019–2023.
5. **Burns, D. P. W., and R. C. Desrosiers.** 1991. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. J. Virol. **65**:1843–1854.
6. **Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biologic features of HIV-1 that correlate with virulence in the host. Science **240**:80–82.
7. **Darby, S. C., R. Doll, S. K. Gill, and P. G. Smith.** 1987. Long-term mortality after a single treatment course with x-rays in patients treated for ankylosing spondylitis. Br. J. Cancer **55**:179–190.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. **12**:387–395.
9. **Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit.** 1991. HIV-1 V3 variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. Virology **180**: 583–590.
10. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155**:279–284.
11. **Goudsmit, J., M. C. Debouck, R. H. Meloen, L. Smit, M. Bakker, D. M. Asher, A. V. Wolff, C. J. Gibbs, and D. Carleton.** 1987. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl. Acad. Sci. USA **85**:4478–4482.
12. **Hahn, B., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232**:1548–1553.
13. **Harper, M. E., L. M. Marselle, R. C. Gallo, and F. Wong-Staal.** 1986. Detection of lymphocytes expressing T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by *in situ* hybridization. Proc. Natl. Acad. Sci. USA **83**:772–776.
14. **Ho, D. D., T. Mougdil, and M. Alam.** 1989. Quantitation of human immunodeficiency virus type 1 in the blood of infected individuals. N. Engl. J. Med. **321**:1621–1625.
15. **Ho, D. D., M. G. Sargadharan, M. S. Hirsch, R. T. Schooley, T. R. Rota, R. C. Kennedy, T. C. Chanh, and V. L. Sato.** 1987. Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. J. Virol. **61**:2024–2028.
16. **Jamieson, B. D., and R. Ahmed.** 1989. T cell memory. Long-term persistence of virus-specific cytotoxic T cells. J. Exp. Med. **169**:1993–2005.
17. **Javaherian, K., A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis, T. Profy, J. R. Rusche, D. P. Bolognesi, S. D. Putney, and J. Matthews.** Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. USA **86**:6768–6772.
18. **Kimura, M.** 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.
19. **Kozarsky, K., M. Penman, L. Basiripour, W. Haseltine, J. Sodroski, and M. Krieger.** 1989. Glycosylation and processing of the human immunodeficiency virus type 1 envelope protein. J. Acquired Immune Defic. Syndr. **2**:163–169.
20. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dressman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249**:932–935.
21. **Li, W.-H., C.-I. Wu, and C.-C. Luo.** 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.
22. **Looney, D. J., A. G. Fisher, S. D. Putney, J. R. Rusche, R. R. Redfield, D. S. Burke, R. C. Gallo, and F. Wong-Staal.** 1988. Type-restricted neutralization of molecular clones of human immunodeficiency virus. Science **241**:357–359.
23. **Ludlam, C. A., J. Tucker, C. M. Steel, R. S. Tedder, R. Cheingsong-Popov, R. A. Weiss, D. B. McClelland, I. Philip, and R. J. Prescott.** 1985. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. Lancet **ii**:233–236.
24. **MacKeating, J. A., J. Gow, J. Goudsmit, L. H. Pearl, C. Mulder, and R. Weiss.** 1989. Characterisation of HIV-1 neutralisation escape mutants. AIDS **3**:777–784.
25. **Meloen, R. H., R. M. Liskamp, and J. Goudsmit.** 1989. Specificity and function of the individual amino acids of an important determinant of HIV-1 that induces neutralising antibody. J. Gen. Virol. **70**:1505–1512.
26. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeld-Manson, B. Asjo, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. Cell **58**:901–910.
27. **Modrow, S., B. H. Hahn, G. M. Shaw, R. C. Gallo, F. Wong-Staal, and H. Wolf.** 1986. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. **61**:570–578.
28. **Neurath, A. R., N. Strick, and E. S. Y. Lee.** 1990. B cell epitope mapping of human immunodeficiency virus envelope glycoproteins with long (19- to 36-residue) synthetic peptides. J. Gen. Virol. **71**:85–95.
29. **Paabo, S., D. M. Irwin, and A. C. Wilson.** 1990. DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265**:4718–4721.
30. **Palker, T. J., T. J. Matthews, M. E. Clark, G. J. Cianciole, R. R. Randall, A. J. Langlois, G. C. White, B. Safai, R. Snyderman, D. P. Bolognesi, and B. F. Haynes.** 1987. A conserved region at the COOH terminus of human immunodeficiency virus gp120 envelope protein contains an immunodominant epitope. Proc. Natl. Acad. Sci. USA **84**:2479–2483.
31. **Psallidopoulos, M. C., S. M. Schnittman, L. M. Thompson III, M. Baseler, A. S. Fauci, H. C. Lane, and N. P. Salzman.** 1989. Integrated proviral human immunodeficiency virus type 1 is present in CD4+ peripheral blood lymphocytes in healthy seropositive individuals. J. Virol. **63**:4626–4631.
32. **Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. L. Lynn,**

**R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1988. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. Proc. Natl. Acad. Sci. USA **85**:3198–3202.

33. **Saag, M. S., B. H. Hahn, J. Gibbons, Y. Li, E. S. Parks, W. P. Parks, and G. M. Shaw.** 1988. Extensive variation of human immunodeficiency virus type-1 *in vivo*. Nature (London) **334**: 440–444.
34. **Schnittman, S. M., H. C. Lane, J. Greenhouse, J. S. Justement, M. Baseler, and A. S. Fauci.** 1990. Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. Proc. Natl. Acad. Sci. USA **87**:6058–6062.
35. **Schnittman, S. M., M. C. Psallidopoulos, H. C. Lane, L. Thompson, M. Baseler, F. Massari, C. H. Fox, N. P. Salzman, and A. S. Fauci.** 1989. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. Science **245**:305–308.
36. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840–5850.
37. **Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J. Virol. **64**:864–872.
38. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and F. Wong-Staal.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLVIII/LAV, the retrovirus of AIDS. Cell **45**:637–648.
39. **Stevenson, M., T. L. Stanwick, M. P. Dempsey, and C. A. Lamonica.** 1990. HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO J. **9**:1551–1560.
40. **Takahashi, H., S. Merli, S. D. Putney, R. Houghten, B. Moss, R. N. Germain, and J. A. Berzofsky.** 1989. A single amino acid interchange yields reciprocal CTL specificities for HIV-1 gp160. Science **246**:118–121.
41. **Tersmette, M., R. E. Y. de Goede, B. J. M. Al, I. N. Winkel, R. A. Gruters, H. T. Cuypers, H. G. Huisman, and F. Miedema.** 1988. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J. Virol. **62**:2026–2032.
42. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses, p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Willerford, D. M., M. J. Gale, R. E. Benveniste, E. A. Clarke, and W. M. Gallatin.** 1990. Simian immunodeficiency virus is restricted to a subset of blood CD4+ lymphocytes that includes memory cells. J. Immunol. **144**:3779–3783.
44. **Willey, R. L., R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, and M. A. Martin.** 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency virus syndrome retrovirus. Proc. Natl. Acad. Sci. USA **83**:5038–5042.
45. **Williams, P., P. Simmonds, P. L. Yap, P. Balfe, J. Bishop, R. Brettle, R. Hague, D. Hargreaves, J. Inglis, A. J. L. Brown, J. Peutherer, S. Rebus, and J. Mok.** 1990. The polymerase chain reaction in the diagnosis of vertically transmitted HIV infection. AIDS **4**:393–398.
46. **Wolfs, T. F. W., J. de Jong, H. van der Berg, J. M. G. H. Tunagel, W. J. A. Krone, and J. Goudsmit.** 1990. Evolution of sequences encoding the principal neutralization epitope of HIV-1 is host-dependent, rapid and continuous. Proc. Natl. Acad. Sci. USA **87**:9928–9942.
47. **Zack, J. A., S. J. Arrigo, S. R. Weitsman, A. S. Go, A. Haislip, and I. S. Y. Chen.** 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell **61**:213–222.
48. **Zagury, D., J. Bernard, R. Leonard, R. Cheynier, M. Feldman, P. S. Sarin, and R. C. Gallo.** 1986. Long term cultures of HTLV-III-infected T cells: a model of cytopathology of T-cell depletion in AIDS. Science **231**:850–853.
49. **Zhang, L. Q., P. Simmonds, C. A. Ludlam, and A. J. L. Brown.** 1991. Detection, quantitation and sequencing of HIV-1 virus from the plasma of seropositive individuals and from factor VIII concentrates. AIDS **5**:675–681.